Top genome sequencing center selects HP Integrity servers

HP success story

Genome Sequencing Center

As a publicly funded center, the more cost-effective it is, the more genomes it can sequence—and the more value it can add. The Genome Sequencing Center is using the HP Integrity server to apply existing tools to larger genomes without rewriting software. And the center can thoroughly test models by making multiple dry runs using various parameters, increasing accuracy and adding more value while still meeting deadlines.

The Genome Sequencing Center (GSC) at Washington University School of Medicine in St. Louis is the third-largest sequencing center in the world. A leader in the large-scale generation and analysis of DNA sequences and a key partner in the Human Genome Project, the publicly funded GSC shares its methodologies and data via free, Web-based repositories.

The GSC constructed the clone map of the human genome and contributed 25% of the finished sequence. Now, under a three-year, $130 million grant from the National Human Genome Research Institute, the GSC is both sequencing other animal genomes to inform the human sequence and advancing the next phase of genomics: the application of information and technology from this project to diagnose, treat, and prevent human diseases.

While assembling genomes of unprecedented size and complexity, GSC researchers are also striving to reduce the time and cost of sequencing. Working with HP and its value-added reseller, St. Louis-based Global Solutions Group, the GSC has met a need common to researchers as they assemble, analyze, and validate ever-larger genomes: at some point, the processes demand higher than normal levels of addressable memory.

Solutions for the adaptive enterprise.

With the HP server, the GSC obtained a tool that it didn't have in the past to move forward with large-scale computational analysis.

As researchers scale computational tools originally written for small sequences to analyze large sequences, their need for addressable memory grows exponentially. Genome centers typically require servers with 64 GB of RAM for completing specific operations in any large-scale assembly and analysis.

At the GSC, they do a lot of computational analyses that require a server with a large volume of RAM. As they scale up using tools written for small sequences, every genome center needs a server with more than 64 GB of RAM for at least one module in any large-scale assembly and analysis.

The absence of such a server caused costly delays and frustration in the productivity-driven GSC environment. A week into a RAM-intensive process, inevitably an operation would crash, forcing researchers to begin the lengthy operation again.

The GSC has chosen the HP Integrity rx5670 server for RAM-dependent processing. Running the Red Hat Enterprise Linux® AS 3 operating system and configured with Intel® Itanium® 2 processors and 96 GB of RAM, the server enables the GSC to complete processes based on RAM-dependent code.

With the HP server, the GSC obtained a tool that it didn't have in the past to move forward with large-scale computational analysis.

Adding more value—faster

As a publicly funded center, the more cost-effective it is, the more genomes it can sequence—and the more value it can add. The GSC is using the HP Integrity server to apply existing tools to larger genomes without rewriting software. And the center can thoroughly test models by making multiple dry runs using various parameters, increasing accuracy and adding more value while still meeting deadlines.

Before the GSC installed the HP Integrity server, its largest single system had just 64 GB of RAM. Finding no Linux-based server on the market with the required RAM, the GSC resorted to crash-prone and costly workarounds.

With the center's previous large-scale server, adding 64 GB of RAM was very expensive, so the researchers subdivided processes into smaller components and adapted existing code with special algorithms to run larger models on Intel processor–based servers.

Aware of the GSC's need for a next-generation server, the Global Solutions Group provided an early evaluation system of the HP Integrity server configured with 96 GB of RAM. The systems integrator also assigned an onsite engineer to help the GSC technical team install and assess the new server.

Taking advantage of the HP Solution Center in Richardson, Texas, one of more than 80 HP Solution Centers around the world, the GSC team was able to securely perform benchmarks and other performance tests on the HP Integrity server. Before purchasing the system, they spent two months pushing the HP server to its limits.

The system stayed up. The HP Integrity server can run larger assemblies on a single server without breaking the process into smaller pieces or writing special algorithms.

Speed and stability
The speed and reliability of the HP Integrity server is based not solely on processing power, but also on the ability of a program to access and use all of the server's 96 GB of RAM. At the core of the HP Integrity server is the Intel Itanium 2 processor, co-developed by HP and Intel. The HP implementation of the Intel Itanium 2 microarchitecture integrates the HP Scalable Processor Chipset. Invented by HP, the HP chipset lowers memory latency and increases bandwidth—unique advantages that enable GSC researchers to solve ever-larger and more complex problems in less time.

HP server outperforms previous large-scale server by more than a factor of two
Moving forward with its next-generation sequencing initiatives, the GSC is deciphering the genetic codes of nonhuman species and, through comparison with the human genome, is assembling sequences that further knowledge of human disease and health.

Subjects of the GSC's large-scale assemblies include the chimpanzee, which, as the closest relative to humans, shares 98% of the same DNA. GSC researcher Shiaw-Pyng Yang is assembling the chimpanzee genome, a vast project that consumes parallel-processing configurations equipped with hundreds of CPUs.

But at a certain stage in the assembly process, the GSC needs a single, large server with a lot of memory and very fast CPU speed to collect and correct all of the data from parallel processing. At this stage, the HP server performs these activities faster than the center's previous large-scale server. With the HP server, researchers can run processes multiple times before releasing the genome for use—and still meet deadlines.

Paving the way to next-generation tools
While a midrange Integrity server suits the GSC's current needs, the Integrity architecture offers systems that bring unprecedented speed to the most demanding tasks.

Warren Gish, Ph.D, a co-developer of BLAST, benchmarked memory-intensive codes on a 64-processor HP Integrity Superdome equipped with the HP-UX 11i operating system and 256 GB of memory.

Using C-based BLAST, researchers detect evolutionary relationships between sequences of different species and, from these relationships, infer the structure and function of previously uncharacterized genes. BLAST is also a tool to match genetic patterns in databases of existing genomes. In next-generation research, such operations require far more memory speed.

In the '80s, when he wrote BLAST, microprocessor speed was not so different from the speed of main memory. Since then, processor speed has ramped up exponentially, far surpassing growth of memory speed. Because most programs like BLAST make high use of memory, having access to large amounts of memory may open the route to much faster implementations of BLAST.

Top genome sequencing center selects HP Integrity servers

HP success story

Customer at a glance

Industry sector: Genomic research
Name: Genome Sequencing Center, Washington University School of Medicine
URL: http://genome.wustl.edu/

Challenge

- Processing of RAM-intensive operations was slow and error-prone

Solution

- HP Integrity rx5670 server with Red Hat Enterprise Linux AS 3 operating system configured with two Intel Itanium 2 processors and 96 GB of RAM
- Consulting by onsite systems engineer from Global Solutions Group
- HP Solution Center provided HP Integrity server for benchmarks

Results

- Faster execution of larger and more complex models
- More thorough testing and value added with less time, effort, and risk
- Productive use of proven tools with no need to modify or rewrite code

To learn more, visit www.hp.com

© 2004 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Intel and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Linux is a U.S. registered trademark of Linus Torvalds.

5982-6149EN, 05/2004

invent